DOCUMENT RESUME

ED 031 276 LI 001 642

Specifications of a Mechanized Center for Information Services for a Public Library Reference Center. Final Report. Part 1, Preliminary Specification: Mechanized Information Services in Public Library Reference Centers.

California Univ., Los Angeles. Inst. of Library Research.

Pub Date 31 Jan 68

Note-49p.; Related documents are LI 001 643, LI 001 644, LI 001 645, and LI 001 646

EDRS Price MF-$0.25 HC-$2.55

Descriptors-Administrative Organization, *Automation, Computers, *Information Centers, *Information Services, Information Storage, Information Systems, Magnetic Tapes, Personnel, *Public Libraries, *State Programs

Identifiers-*California

This document presents preliminary specifications for a library-based Center for Information Services (CIS). Four sets of issues are covered: (1) data base inventory, providing a listing of magnetic tape data bases now available from national sources or soon to be so; (2) administrative issues, including the organization of the CIS within the library, its administrative relationship to other activities, its staffing, its method of operation, and its service load; (3) hardware issues, including library/CIS computer configuration and its requirements for space; (4) software issues, including the requirements for generalized programs to handle file management and search, reference retrieval, and text processing. (Author)

LI 001 642

ED031276

PRELIMINARY SPECIFICATION: MECHANIZED INFORMATION SERVICES IN PUBLIC LIBRARY REFERENCE CENTERS

Part 1 of the Final Report on

Specifications of a Mechanized
Center for Information Services
For a Public Library Reference Center

STSA Grant A 6 11-66

31 January 1968

Institute of Library Research
University of California
Los Angeles, California

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TABLE OF CONTENTS

ABSTRACT. i

I. INTRODUCTION. 1

II. AVAILABLE DATA BASES. 7

III. ADMINISTRATIVE ISSUES 18

Administrative Organization 18
Library Internal Organization 18
Staffing. 18
CIS Method of Operation 24

IV. COMPUTER CONFIGURATION. 27

The Proposed Library Computer System. 27
Central Processing Unit 27
Input/Output Units. 30
Storage Devices 30
Teleprocessing Components 31
Physical Facilities 31

V. SOFTWARE FOR THE CIS SYSTEM 36

Specifications for CIS File Management Software 38
File Definition 39
File Organization Concepts. 40
File Search Concepts. 40
System Monitoring 41
Systems Functions 41
System Operation. 42
Specifications for CIS Reference Retrieval Software . . 43
File Definition 43
File Organization Concepts. 44
File Search Concepts. 44
System Monitoring 45
System Functions. 45
Specifications for CIS Text Processing Software 45
File Definition 45
File Organization Concepts. 45
File Search Concepts. 46
System Functions. 46

ABSTRACT

Part 1 of the Final Report on Specifications of a Mechanized Center for Information Services in a Public Library Reference Center presents preliminary specifications for a library-based "Center for Information Services". Four sets of issues are covered:

1. Data base inventory, providing a listing of magnetic tape data bases now available from national sources or soon to be so.

2. Administrative issues, including the organization of the CIS within the library, its administrative relationship to other activities, its staffing, its method of operation, and its service load.

3. Hardware issues, including library/CIS computer configuration and its requirements for space.

4. Software issues, including the requirements for generalized programs to handle file management and search, reference retrieval, and text processing.

I. INTRODUCTION

This report presents specifications for the development of mechanized information services in reference centers for the State of California, with special emphasis on service to business and industry. It is the first part of the final report on a study sponsored by the U. S. Department of Commerce under STSA (the State Technical Services Act of 1965).

The theme of library service in California (and elsewhere) is that of expanding scope. If California's productive economy and rich cultural life are to be maintained, then access to book and other library materials must be increased. Unless there are sound local libraries backed up by a means to draw on distant library resources, severe handicaps are imposed upon every level of society: the pre-school child for whom the library provides an introduction to the world; the beginner reader with his insatiable curiosity; the student and his need for reference materials; the adult citizen and his need for information on family, social and political life; the research scientist and the technical specialist and their needs for specialized information.

But in addition, today's library is called on to serve even wider needs for library service and is paying increasing attention to "information" services. Such information services include, in addition to library service: (1) information analysis; (2) publication, announcement, and distribution; (3) information generation and usage. The public library is assuming more responsibility in all of these areas, and eventually

should serve as an agency for acquisition of data not previously considered within its scope, including in particular, machine-readable computer data. It will also serve as a point for access to state, national, and even international resources through networks of various kinds. Perhaps most important, the library can provide a point of assistance in the use of these new forms of data.

This implies a need for specification of a mechanized "Center for Information Services" to be installed in the Public Library Systems of the country to meet the requirements for information services under the State Technical Services Act of 1965.

This Act arose out of demands to speed the spread of technology developed under government sponsored projects into civilian industry. It has as its purpose the diffusion and application of science and technology in business, commerce, and industry. In addition to educational functions, the Act defines "technical services" to include:

(a) Preparing and disseminating technical information in a variety of forms, specifically including computer tapes and microforms;

(b) Establishing technical information centers to carry out that preparation and dissemination; and

(c) Provide reference centers to identify sources of expertise.

The Act thus clearly defines a set of library activities. It requires the acquisition, storage, and distribution of recorded data, including reports, abstracts, and reviews in the form of printed documents as well as mechanized media such as magnetic tapes and microforms. It specifically calls for establishment of technical information centers which must include the ability to utilize these data forms.

The development of these centers must be directed toward their becoming an ongoing, operational system: i.e., they must provide day-to-day information services. Furthermore, their services must be immediately accessible to even the smallest of businesses in local communities throughout the State. The administration will therefore require a high level of experience in providing library types of services.

The kinds of activities discussed above are currently provided by the complex system of public libraries of the country and in particular by the State Libraries. This system of libraries is therefore administratively well suited for the operation of the centers called for in the State Technical Service Act--once they have been developed, and provided the conditions necessary to introducing such centers into the public library system have been considered in their planning.

During 1967, the Institute of Library Research of the University of California studied Mechanized Information Services in Library Reference Centers.

The study was concerned with library services for handling media such as magnetic tape. Since these machine-readable data bases have been developed for a variety of purposes outside those normally considered within the scope of the library, several problems are faced by the library in extending its scope to include acquiring such media, cataloging them, and providing "information services" based on them.

Some of the issues relate to the content: What kinds of material should the library acquire? Some of them concern library processes: How do we catalog magnetic tape materials? Some of the problems are technological: How do we provide man-machine communication? Some of them are administrative:

How do we finance information services? How do we fit them within the traditional library structure?

The interest in such services is a natural result of the great number of efforts to develop mechanized information services and produce national information networks with a high level of mechanization. Part II of this report therefore provides a context within which to view the development of the State library network. Part III provides a quantitative picture of the present state of the network.

Some problems represent essentially policy issues, since there is simply not enough data to resolve them on an objective, factual basis:

1. Is it worthwhile to provide mechanized information services to the business and industrial community?

2. Should the public library be regarded as the appropriate agency for such services?

3. How should the public library proceed in relation to efforts in development at other libraries and at a national level?

Part IV of this report describes the approach taken to study of these policy problems and summarizes the results.

Other problems are essentially technical, relating to the characteristics of mechanized data bases and the requirements for programs to process them. This part of the report summarizes the results (the study of these technical issues).

The addition of machine readable media to the library's collection will require additions of staff, changes in internal administrative organization, and the formalization of relationships with other activities. The preliminary specifications therefore present an organization chart in which CIS Departments, reporting to an Assistant Librarian for Mechanized Services, provide coordination and liaison of CIS activities and operation and system support of its computer installation. Staffing requirements are enumerated.

Although exact specifications for a computer facility will almost certainly be changed before installation a reasonable minimal system is presented which will provide both on-line and batch processing capabilities for the library's computer oriented services. It is designed to serve both Information Services and production processing.

The success of library services ultimately will depend on effective programming. Study of the alternatives for handling files produced for many differing original purposes has led us to specify that CIS software should be "generalized" and able to handle a wide variety of formats. The specifications call for three separate modules. The first, CISFMS (Center for Information Services File Management Software), is a general purpose system for normal file maintenance, servicing requests for simple, field-structured searches. It quickly puts acquired data bases into service with minimum demands on both programmer and user. For processing more complex requests, a second module known as the CISRRS (CIS Reference Retrieval Software) is specified. It will search data bases which involve the use of subject descriptions and in which at least two files are interactive (e.g., master files and index files). It provides for more sophisticated processing where repeated field data are involved. The third module, CISTPS, (CIS Text Processing Software) is designed around the particular needs of generalized text processing.

In summary, the concept of a Center for Information Services is engendered by the developments of modern information technology. Organizationally the Center is viewed as an administrative part of the library. Physically it is viewed as a storage and processing facility. It will provide a supplement to the media and method of operation of the usual library. It must have an ability to deal with a wide variety of

data bases and programs. It will require new policies and procedures, new relations to other organizations, and new means of cooperation with other centers. The system must be operational, general purpose, adaptable, replicative, and designed to encourage easy use.

II. AVAILABLE DATA BASES

This survey is based largely upon information compiled for the Institute of Library Research by Informatics, Inc., in their report, In Specification of a Center for Information Services; Appendix A: Descriptions of Data Bases, Sherman Oaks, California, 1967. The listing here emphasizes reference data bases and does not claim to be exhaustive even in that coverage; however, it is indicative of the growing variety and number of magnetic tape files in existence, of a type which might be utilized in a Center for Information Services in the University Library. It reflects, for the most part, projects undertaken on a large national scale, or to serve the needs of particular organizations. A National Science Foundation publication, Nonconventional Scientific and Technical Information Systems in Current Use, No. 4, December 1966, contains an additional listing of more than one hundred computer-based information retrieval systems which utilize reference data bases. In almost all cases, the primary storage medium is magnetic tape.

There are also increasingly large numbers of machine readable files, many of them available at nominal charge, being created by individuals or by small groups in industrial organizations, or within university departments. A number of these (emphasizing text data bases) are noted in compilations such as Literary Works in Machine Readable Form[1], and

---

[1] Carlson, Gary, Literary Works in Machine Readable Form, by Dr. Gary Carlson, Director, Computer Research Center, Brigham Young University, Provo, Utah. July 1965. (This list is updated in the January 1967, issue of Computer and the Humanities.

Computerized Research in the Humanities: A Survey[1]. The Council on Social Science Data Archives has published a brochure, Social Science Data Archives in the United States, 1967[2], which lists and describes files covering a wide spectrum of subject matter (emphasizing numerical data bases), many of which are available from sponsoring institutions.

A commerical publication, Directory of Computerized Information in Science and Technology, 1967[3], is scheduled for issue in Spring 1968. Other directories, covering computerized information in Medicine, the Humanities, and the Social Sciences are planned. These will be published as part of an "International Information Network Series" and will serve to bring the existence of many more machine readable files to current awareness.

In the following pages, the address and director of the creating agency is listed for each of a variety of data bases now available or soon to be so.

Several overall observations about data bases can be made from an examination of this listing:

Many of the files were created for specific purposes and were tailored to meet the special needs of the parent organization. Therefore, they have been designed without regard to a capability for easy readability

---

[1]Bowles, Edmund A., "Computerized Research in the Humanities: A Survey". ACLS Newsletter Special Supplement (June 1968) 1-49.

[2]"Social Science Data Archives in the United States, 1967". Council on Social Science Data Archives, New York, New York.

[3]Directory of Computerized Information in Science and Technology Part I, 1968. New York, Science Associates/International.

for other purposes. Documentation in such cases is frequently poor and incomplete, and cooperation is apt to be uncertain or unenthusiastic.

On the other hand, some organizations (both profit as well as non-profit) are in the business of maintaining data bases and providing a variety of services--searching, preparing reports, copying files, and producing extracts or sub-files. These data bases are generally, but not always, easy to read and well documented, and are usually furnished with computer programs to read, search, and otherwise process the data involved.

The majority of organizations surveyed use IBM equipment, particularly 1401/1410 systems. Most, if not all of these, are converting to 360 systems. The use of tapes is still dominant, the trend to greater use of discs being, at the moment, quite small.

From a file management point of view, most of the existing data bases have simple, hierarchically arranged, field structures. Many have variable length records. Record formats (fixed or variable), from one file to another, are virtually unrelated. It is evident that translation or transliteration to a common format is nearly impossible, and custom programming a complete system for each data base is far too expensive. The maintenance and use of programs written by sponsoring organizations appears to be cumbersome and impractical (for example, there are 15 programs involved in the American Petroleum Institute system), and the incompatibility of software systems adds to the difficulty.

American Bibliographical Center
2010 Alameda Padre Serra
Santa Barbara, California 93103
Director: Dr. Eric Boehm
Historical Abstracts

American Chemical Society
Publications Department
1155 Sixteenth Street, N.W.
Washington, D. C. 20036
Director of Business Operations: Joseph H. Kuney
Journal of Chemical Documentation
Journal of Chemical Engineering Data

American Geological Institute
1444 North Street, N.W.
Washington, D. C. 20050
Geoscience Abstracts
Bibliography and Index of Geology, Exclusive of North America

American Petroleum Institute
Division of Refining
Central Abstracting and Indexing Service
555 Madison Avenue
New York, New York 10022
Manager: Mr. Everette H. Brenner
Petroleum Abstracts

American Society for Metals
ASM Documentation Service
Metals Park, Ohio 44073
Director: Norman E. Cottrell
Associate Director: Mrs. Marjorie Hyslop
Review of Metal Literature

Applied Mechanics Review
Southwest Research Institute
8500 Culebra Road
San Antonio, Texas
Director: Mr. Stephen Juhasz
Applied Mechanics Review

Atomic Energy Commission
Atomic and Molecular Processes Information Center
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
Director: C. F. Barrett
Atomic and Molecular Processes Information

Atomic Energy Commission
Division of Technical Information Extension
Post Office Box 62
Oak Ridge, Tennessee 37831
Chief, Computer Operations: Joel S. O'Connor
Nuclear Science Abstracts

Biosciences Information Service, of
Biological Abstracts
3815 Walnut Street
Philadelphia, Pennsylvania 19104
Director: Phyllis V. Parkins
Assistant Director for Systems Development:
Miss Louise Schultz
AUTHOR index
BASIC (Biological Abstracts Subjects in Context)
CROSS (Computerized Rearrangement of Special Subjects)
BIOSYSTEMATIC

U. S. Department of the Interior
Bonneville Power Administration
Portland 8, Oregon
System Engineer: Val Lava
Electrical Engineering Abstracts

R. R. Bowker Company
1180 Avenue of the Americas
New York, New York 10036
Book Editorial Department: Mr. John N. Berry, III
American Book Publishing Record
Forthcoming Books
Publisher's Weekly
Paperbound Books in Print
Subject Guide to Books in Print
Children's Books for Schools and Libraries

U. S. Department of Commerce
Bureau of the Census
Washington, D. C. 20233
Director: A. Ross Eckler
Available tape files cover population, housing,
agriculture, business, foreign trade, etc.

Bureau of Labor Statistics
U. S. Department of Labor
Washington, D. C. 20210
Survey of Industry Labor Turnover
National Survey of Scientific and Technical Personnel in Industry
Survey of Industry Employment Payroll and Hours
Survey of Industry Employment, Worker Earnings and Hours of Work for States and Areas
Estimates of Labor Force Characteristics from Current Population Survey
Survey of Consumer Expenditure
Occupational Outlook Matrix
State and Area Employment and Earnings
Industry Sector Price Indexes

University of Saskatchewan
Regina Campus
Regina, Saskatchewan
Canada
Canada News Index (planned)

Canada--Department of Forestry and Rural Development
Geo-Information System of the Canada Land Inventory
Ottawa, Canada

Chemical Abstracts Service
2540 Olentangy River Road
Columbus, Ohio
Director: Dale B. Baker
Manager, Subscriber Information Service: Mr. Elden G. Johnson
Chemical Titles
CBAC (Chemical and Biological Activities)
POST (Polymer Science and Technology)
Chemical Compound Registry

Clearinghouse for Federal Scientific and Technical Information
5825 Port Royal Road
Springfield, Virginia 22151
Director: Bernard Fry
Assistant Director, Systems: Peter F. Urbach
U. S. Government Research and Development Reports

Johns Hopkins University
Baltimore, Maryland 21205
Communications in Behavioral Biology

Computer Software Management and
Information Center (COSMIC)
Computer Center (C-B)
University of Georgia
Athens, Georgia 30601
(no publication)

Direct Access to Reference Information,
A Xerox Service (DATRIX)
University Microfilms Library Services
Xerox Corporation
Ann Arbor, Michigan 48106
(searches on request)

U. S. Office of Education
Educational Research Information Center (ERIC)
400 Maryland Avenue, S.W.
Washington, D. C.
Director, Division of Information Technology and
Dissemination: L. G. Burchinal
RIE (Research in Education)

Engineering Index
345 East 47th Street
New York, New York 10007
Assistant General Manager: Mr. Michael Tomaino
Electrical/Electronics Engineering

Engineers Joint Council
345 East 47th Street
New York, New York 10007
Mr. Frank Speight
Thesaurus of Engineering Terms

Frost & Sullivan, Inc.
179 Broadway
New York, New York 10007
Mr. Daniel M. Sullivan
DM (2) (Defense Market Measures)

General Electric Corporation
Flight Propulsion Division
Cincinatti, Ohio 45215
Manager, Information Systems: George Carr

Harvard University
Vision Information Center (NINDB)
Countway Library of Medicine
Boston, Massachusetts 02115
(no publication)

Health Law Center
Graduate School of Public Health
University of Pittsburgh
Pittsburgh 13, Pennsylvania
Assistant Director: Eric W. Springer
Total State Statutes
Health Statutes
U. S. Appropriation Acts
Internal Revenue Regulations
etc.

Johns Hopkins University
Information Center for Hearing and Speech
and Disorders of Human Communication (NINDB)
Baltimore, Maryland 21205
(no publication)

Hughes Aircraft Company
Los Angeles, California
Electronic Properties Information Center (EPIC)

Institute for Scientific Information
325 Chestnut Street
Philadelphia, Pennsylvania
Director: Dr. Eugene Garfield
Director of Research: Dr. Irving H. Sher
Science Citation Index

International Labour Office
Central Library and Documentation Branch
Integrated Scientific Information Service (ISIS)
Geneva
Weekly Bulletin

Library of Congress
Information Systems Office
1st Street and Independence Avenue, S.E.
Washington, D. C.
Director, MARC Pilot Project: Mrs. Henriette Avram
Project MARC (Machine Readable Catalog)

Library of Congress
Card Division
Building 159 Navy Yard Annex
Washington, D. C. 20541
Chief, Card Division: Alpheus L. Walter
Subject Headings

National Aeronautics and Space Association
400 Maryland Avenue, S. W.
Washington, D. C. 20202
Director, Scientific and Technical Information
Division: John F. Stearns
Scientific and Aerospace Reports
International Aerospace Abstracts

National Bureau of Standards
U. S. Department of Commerce
Office of Technical Information and Publications
Washington, D. C. 20234
Chief: W. R. Tilley
Index of Government Sponsored Computer Projects
National Standard Reference Data System
Crystal Data Determinative Tables

National Council on Crime and Delinquency
Information Center on Crime and Delinquency
44 East 23rd Street
New York, New York
International Bibliography on Crime and
Delinquency

National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland
Associate Director for Intra-Mural Programs:
Joseph Leiter, Ph.D.
MEDLARS CCF (Condensed Citation File)

New York Times Index
Times Square
New York, New York 10036
New York Times Index

Ontario Institute for Studies in Education
Toronto 5, Ontario
Canada
Carnegie Human Resources Data Bank
(publishes various bulletins and searches on request)

PANDEX
American Management Association Building
135 West 50th Street
New York, New York 10020
PANDEX (printed, microfiche, magnetic tape)

Parkinson's Disease Information and
Research Center (NINDB)
Columbia University
New York, New York 10032

Linguistics Department
Rand Corporation
1700 Main Street
Santa Monica, California 90406
Bibliography of Computational Linguistics
(various textual files, including 30 million words of Russian text)

Science Information Exchange
Smithsonian Institution
209 Madison National Bank Building
1730 M Street, N. W.
Washington, D. C. 20036
Director: Monroe E. Freeman, Ph.D.
The Grant Master File

Stanford University Libraries
Stanford, California 94036
Computer Produced Catalog

Brain Information Service (NINDB)
Biomedical Library
53-233 Health Sciences Center
University of California
Los Angeles, California 90024

Department of Political Science
Statistical Laboratory
4343 Social Science Building
University of California
Los Angeles, California 90024
Director: Dwaine Marvick
"POLCEN" (Political Census)

University of Southern California--McGraw-Hill
Division of Cinema
University of Southern California
Los Angeles, California
Director: Glen McMurry
National Information Center for Educational Media (NICEM)

Project URBANDOC
The City University of New York
33 West 42nd Street
New York, New York 10036
Director: Mrs. Vivian Sessions
URBANDOC

United States Department of Agriculture
Washington, D. C.
Current Research Information Systems (CRIS)
(searches on request)
Pesticides Information Center
(will output special bibliographies, also
search on request)
Bibliography of Agriculture

III. ADMINISTRATIVE ISSUES

A Center for Information Services will operate as the administrative agency for coordination of the acquisition, cataloging, storage, and processing of machine processible data. It is therefore necessary to establish at least a preliminary definition of its administrative structure, its staffing, and its method of operation.

ADMINISTRATIVE ORGANIZATION

Library Internal Organization

Operation of the Center for Information Services will involve four major groups in the library: the Library Technical Services Department, for acquisitions and cataloging; the Library Reference Department, for service and public relations; the Library Data Processing Department, for computer operations; and the Library Systems Department, for development and maintenance. A "CIS Administrative Department" will coordinate their activities, provide special expertise in information handling as necessary, and serve as liaison with activities at other libraries in the network.

Figure 1 is a schematic organization chart in which the role of each department in the operation of a CIS has been highlighted.

Staffing

The CIS operation will require addition of new staff as well as special training for existing staff of the library. However, there are some particularly difficult staffing problems. The following paragraphs enumerate the kinds and numbers of personnel required, and in each case

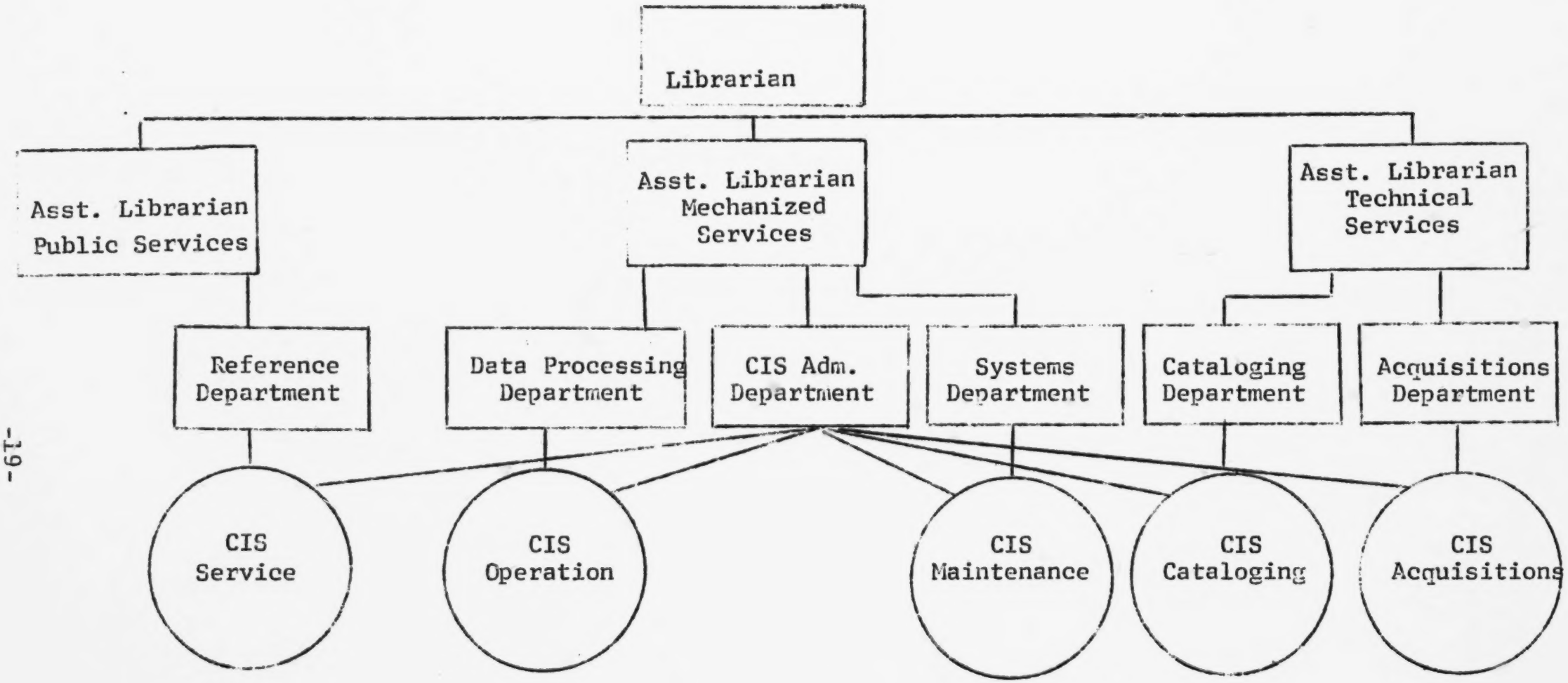

Librarian
Asst. Librarian Public Services
Asst. Librarian Mechanized Services
Asst. Librarian Technical Services
Reference Department
Data Processing Department
CIS Adm. Department
Systems Department
Cataloging Department
Acquisitions Department
CIS Service
CIS Operation
CIS Maintenance
CIS Cataloging
CIS Acquisitions

FIGURE 1

estimate the salary level which the position calls for in terms of an existing library salary structure. Unfortunately, there are serious inconsistencies between those salaries and the competition for the limited number of people who combine knowledge of data processing and information retrieval with knowledge of libraries. The problems raised are particularly acute for the position of "Assistant Librarian for Mechanized Services" and those of the three department supervisors.

It is possible that in order to attract personnel with the required competence, it will be necessary to depart radically from the existing salary scales. On the other hand, that would raise problems in the working relationship between people in library positions of equivalent responsibility but with disparate salaries. Because those problems ultimately could destroy the effectiveness of the CIS, it has been assumed that salary scales consistent with others in the library will be used. This means that personnel must be found among those with less experience but real capability.

A key person in CIS operations is the Assistant Librarian for Mechanized Services. Within the general guide lines established by the Librarian, he is responsible for the analysis and design of the library's information system and for administration of the professional staff required for such work. He analyzes prospective projects to insure that all sources of data pertinent to the program have been identified. He evaluates existing information services and those proposed for the future with regard to user needs, efficiency of equipment, and methods of operation. He applies detailed technical knowledge of both computer based and manual information storage and retrieval systems in such evaluation. He prepares specifications for such services, including relating them with various existing programs.

He communicates results as necessary to carry out liaison with organizations, agencies, and individuals, on campus as well as off campus. A salary within the range of $14,000 to $16,000 should be planned on.

Falling under the direction of the Assistant Librarian for Mechanized Services are those aspects of the CIS operation which involve the mechanized equipment and its utilization. Specifically, he is responsible for the CIS Administrative Department, the Data Processing Department, and the Systems Department.

Since the CIS Administrative Department will provide the special expertise in information handling in support of the other departments of the library, its primary staffing needs are for "information specialists".

They will function under the direction of the Supervisor of the CIS Administrative Department (an Associate Information Specialist with a salary of $12,000 to $13,000). He plans, organizes, and coordinates the activities of the other departments of the library to assure the successful operation of the CIS as a program entity. He provides liaison with other libraries with respect to mechanized information services. He assigns information specialists under his direction to assist in determining requirements for acquisitions, in cataloging and describing acquisitions properly, in phrasing of requests for service, and in scheduling the processing of the files.

His staff consists of two or more Assistant Information Specialists (with salaries of about $10,000), who serve as the means for communication between libraries and the specialized technical data files and reference files of the CIS. They evaluate index data to insure complete accuracy in description of material and appropriate depth of indexing for value in later retrieval. They assist in determining needs for information in formulating requests, and in analyzing requests, and in analyzing retrieved information

for presentation to the user. They ensure appropriate dissemination of incoming information to users. They have sufficient technical knowledge of CIS information storage and retrieval systems to use them effectively.

The Data Processing Department provides for the management and operation of the library's computer-related equipment facilities, including not only the computer itself but peripheral equipment elsewhere in the library.

It operates under the direction of the Supervisor of the Data Processing Department (with a salary of $12,000 to $13,000). He plans, organizes and controls the operation of the computer and peripheral data processing equipment, and is in full charge of all library computing equipment operations. He establishes detailed schedules for the utilization of all equipment to obtain maximum usage. He assigns personnel under his direction to the various operations and instructs them where necessary so they are trained to perform assigned duties in accordance with established methods and procedures. He provides technical liaison with computer facilities outside the library to coordinate activities. He reviews equipment logs and reports on equipment operation efficiency. He must be familiar at the working level with all phases of the operation, and should have a knowledge of computer programming sufficient to diagnose malfunctions as due to operation, equipment, or programming.

He will require the assistance of a Lead Computer Operator for each shift (with a salary of about $10,000). They should have technical knowledge of computer operations comparable to a Senior Computer Operator (see below) and also supervisory capability for instructing, assigning, directing, and checking the work of the other computer operators, including the seniors. They assist in the scheduling of the operations and the assignment of personnel to the various items of equipment required for computer functions.

They may act as shift supervisors in the absence of the Department Supervisor. The staff should include a Senior Computer Operator for each shift (with a salary of about $8,000), who should be competent to work at all phases of computer operation with very little assistance. Other personnel--a total of perhaps ten for a two shift operation--should include junior or trainee computer operators (3; 2 for a day shift, 1 for a second shift), a clerical staff to receive requests for use of computer and organize them for processing according to predetermined rules, and key-punchers. Their salaries are in the range of $5,000 to $7,000 each.

Under the general direction of the Assistant Librarian for Information Services, the Systems Department is responsible for analysis and design of data processing and information handling systems for the library. It functions under the direction of the Supervisor of the Systems Department who is an Associate Information Systems Specialist (with a salary of $12,000 to $13,000). He is responsible for direct supervision of analysts and programmers including outlining detailed procedures to be followed. He works closely with librarians and other library personnel in the definition of their specific requirements.

Under his direction is a staff of Assistant Information Specialists (with salaries of $8,000 to $12,000) who are capable of one or more of the following tasks: analysis of information handling functions and in the development of general system design, application of analytical techniques to the study and evaluation of both existing systems and alternative ones, application of existing systems and procedures to assigned tasks, conversion of existing operations to new ones, preparation of detailed manuals for operation. Also under his direction is a staff of Programmers (with salaries in the range of $8,000 to $12,000), responsible for the actual work of programming the computer.

In other library departments there is no need for additional or different personnel, as such. There is, however, a real need for education of the present staff in the particular problems of mechanized services and the methods for solution.

CIS METHOD OF OPERATION

It is presently visualized that a patron seeking to use the CIS will present a request to a local library, who will determine whether the request for information can best be handled by local resources and conventional procedures (such as consultation of the card catalog, a bibliography, or other reference tool) or by reference to the CIS. When the librarian recommends use of mechanized data, he will help the user formulate his request.

It is expected that the CIS itself will be limited by the small storage capacity, relatively slow processing speed, and moderate peripheral equipment of the small computer it uses. It will therefore operate as a batch processing system. Requests for CIS searches will either be accumulated on-line or written and forwarded on a daily basis to the library computer. Searches will be run against the various files on a scheduled basis. Output will be provided in printed form or on other media as requested, or will be transmitted to the campus computer.

The lack of experience with systems of the CIS type make it impossible with any real confidence even to estimate the number of requests to be expected. The projections of workload have been based on a figure of five files per day to be scheduled for search, with the number of requests processed against each varying from one to many.

Over the next five year period, it is expected that the CIS will acquire at least twenty data bases (each of which may involve three or four files of at least six tapes each per year) for a total of about 2,000 tapes.

The processing time per request depends upon the degree of batching, the basis for scheduling of specific files or portions of files, the tape running time and the number of tapes involved, the internal (CPU) processing time, etc. Maximum file size has a direct bearing on file search time. Most data bases are small (a few reels of magnetic tape), but one data base in the survey consists of 50 reels. The maximum allowable record size encountered in the data bases examined so far is about 54,000 characters. This is an extreme however; more typically, allowable records are limited to about 2,000-3,000 characters. (These are maximum allowable sizes for variable length records; the actual limits on size of large records is unknown).

As an initial approximation, we estimate an average of four tapes per portion of a file scheduled to be processed on a given day, taking a total of one hour per file. Thus, processing the anticipated five files per day, CIS would utilize about 30% of the two-shift capacity of the library's computer.

Storage and handling of magnetic tape files by library personnel at a central location should minimize loss or destruction of data. Duplicate tapes for outside processing would be supplied on a regular chargeout basis. Particular care must be taken to guarantee that duplication and dissemination of tape files does not infringe upon any copyright of the organizations which originally issued them.

In this method of operation, we have emphasized the desirability of controlled access. We feel that it will be some while before access to

mechanized data will be simplified so that any but specially trained personnel will be able to use it effectively. This is not to underestimate the value of mechanized retrieval services, since we feel they will become an essential part of library operation, but simply to emphasize what we feel is a significant point in the economic utilization of such services. Mechanized information services are of primary value as an aid to the professional, trained in information services in general or in their application to a particular area.

IV. COMPUTER CONFIGURATION

Defining the detailed configuration for a computer installation to be operative in say 1972, for workloads largely unknown (or undeterminable at this time), is at best an iterative process. Any initial plan for configuration is almost certain to be modified to take into account changing workloads, unforeseen requirements, new software, etc. However, the basic configuration described in this section (an IBM System/360 Model 30 with 64K core, 2311 disc files, and tape drives) is considered minimum from a CIS point of view at this stage (Phase I) of the CIS project.

THE PROPOSED LIBRARY COMPUTER SYSTEM

Although the system described below is a relatively small one for the variety of applications planned, it can accomplish both on-line and batch processing, and seems adequate to provide both sufficient time and hardware capabilities for the library's computer-oriented services. The configuration is summarized in Figures and .

Central Processing Unit

The system is built around an IBM 2030-F central processing unit with a 64,000 character core memory. The required features to be added to the processing unit are: decimal arithmetic capability for computational purposes; an internal timer, which will be indispensable for the proposed monitor system; a selector channel to handle certain

FIGURE 2

LIBRARY COMPUTER SYSTEM SCHEMATIC

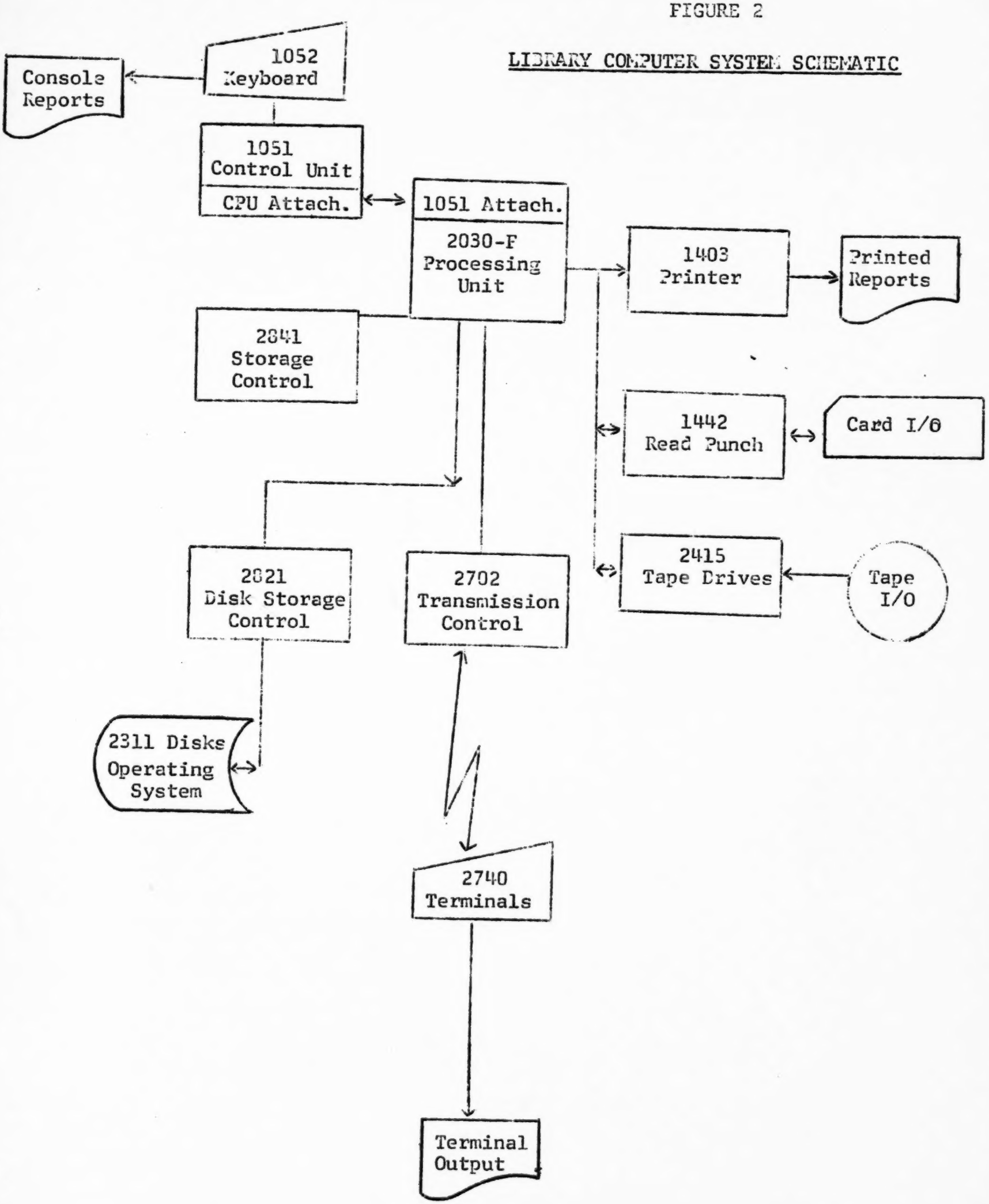


FIGURE 3

Tentative CIS Configuration

IBM 360/30

2030 F		Processing unit (64K)	$178,390
	3237	Decimal arithmetic	1,000
	4760	Internal timer	1,940
	6960	Selector channel	8,290
	7520	Storage protect	5,820
	7915	1051 Attachment	4,005
	4427	Floating point (PL/1)	1,940
1051 N1		Control unti	3,050
	3130	CPU attachment	500
1052 6		Printer keyboard	2,725
1442 N1		Card read punch	26,250
1403 2		600 lpm printer	34,000
	8641	Universal character set	450
	4740	Interchangeable chain cartridge adapter	3,125
2821 2		Control unti	27,940
2311 1		Disk storage drive (4 @ $25,510)	102,040
	1316	Disk Pack (8 @ $490)	3,920
2841 1		Storage control	26,430
	4385	Filescan	1,360
	6118	Record overflow	400
2702		Transmission control	39,580
	1065	Additional selective speed	700
	4615	Terminal control Type I	1,575
	4616	Terminal control Type II	1,575
	4612	Line adapter (4 @ $1,385)	5,560
2740		Communication terminal (4 @ $3,100)	12,400
	3255	Dial up (4 @ $135)	540
	4790	Line adapter (4 @ $450)	1,800
2415 2		Tape drive (800 bpi, 14 kb)	57,040
			$554,345

input and output functions; a storage protect feature to provide safety in multi-terminal processing; a keyboard terminal attachment for interface with the console; and the capability to handle floating point arithmetic, required for implementation of the PL/1 programming language.

The memory of the proposed system may be reduceable to 32K rather than th 64K proposed. Part of this issue relates to the operating system to be used (as discussed in the next section), since DOS fits into 32K but OS/360 requires at least 64K. In addition to requiring less core, DOS has the advantage of being faster in some cases; OS, on the other hand, provides vastly more data management service and is, on the whole, considerably more sophisticated and flexible than DOS.

Input/Output Units

Input/Output devices include a console printer/keyboard; a printer capable of producing 600 lines per minute with a universal character set and including features to enable use of the upper- and lower-case print chain; a dual magnetic tape drive; and a card reader-punch.

Storage Devices

The storage control unit of the proposed system includes a file-scan device for additional input-output protection and a record overflow indicator. The system includes four disk drives and eight disk packs for auxiliary storage.

Teleprocessing Components

The remainder of the configuration provides the teleprocessing capabilities of the system and is based on keyboard communication terminals, as well as the data collection terminals for use in circulation control. The new terminals will have dial-up and line adapter features. The transmission control device attached to the processing unit will handle both types of remote terminals.

The proposed system is a minimal but adequate computer configuration for library processing and CIS jobs. It will be possible to connect it to a larger computer for more powerful processing, using the transmission control devices of the two systems.

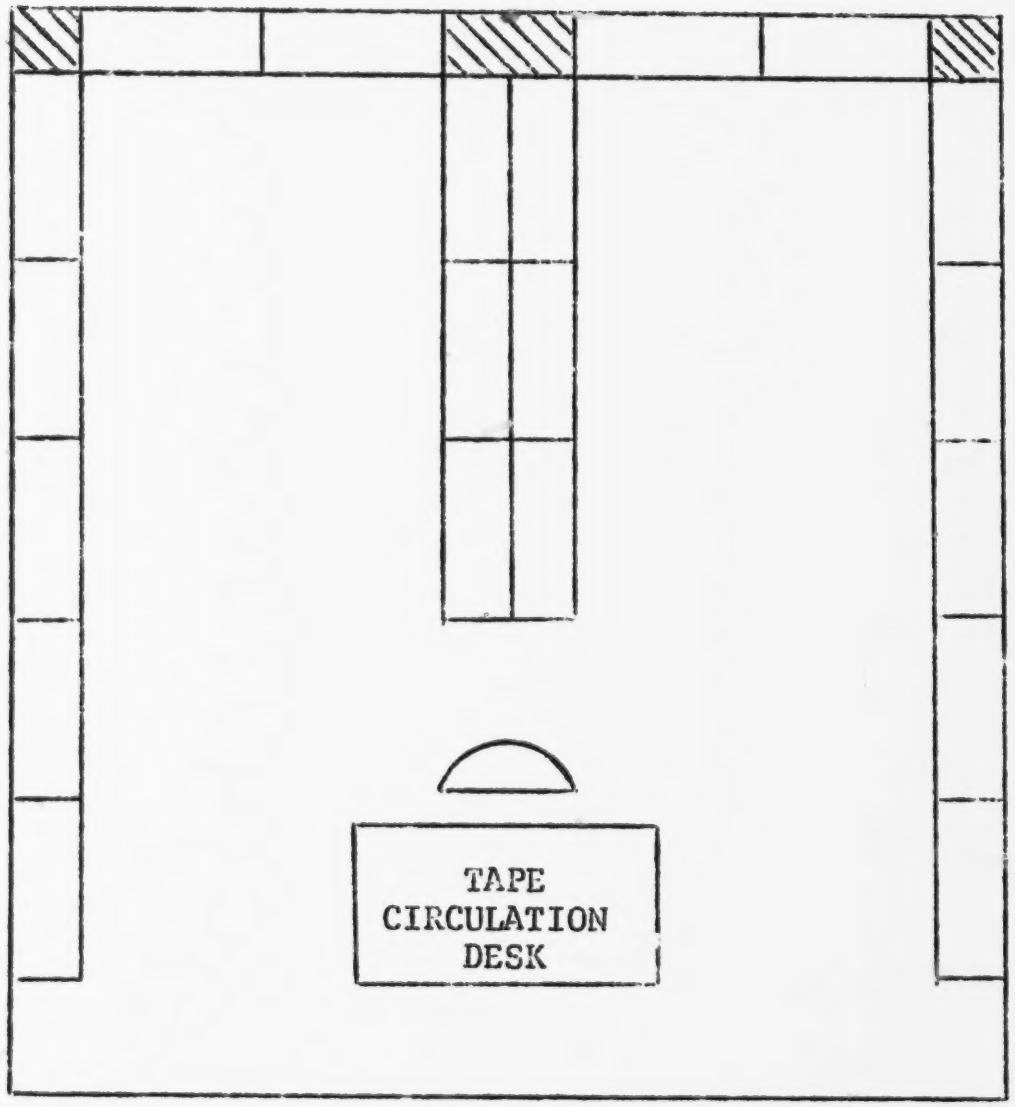
PHYSICAL FACILITIES

At the architectural level the use of computers means a new look at library layout, since the effects of automation can radically change organizational relationships in library technical services and the flow of information and material. At the engineering level, it means concern with environmental control, with needs for cabling, with structural planning for equipment, with lighting for consoles and microform reading, with acoustics of input and output devices such as typewriters and printers.

Provision must be made for the central processing facility itself. A preliminary space allocation is as follows:

(A) For the central processing facility itself--1,000 square feet (see Figure 4).

(B) For immediately adjacent service area for storage of spare parts and text equipment--100 square feet.



Notes

--Each tape bay shown in 36 inches wide and has six shelves. The total capacity of a bay is 120 tapes. There are 20 bays shown with a total capacity of 120 times 20, or 2,400 tapes.

--The area included in the space plan is 17 feet by 18 feet, or 306 square feet.

FIGURE 4 : Tape Storage Space Plan

(C) For storage of tapes, discs, and other forms of mechanized storage--300 square feet adjacent to the central processing facility (see Figure).

(D) For storage of cards, forms, and other supplies--200 square feet (located away from central processing facility, but convenient to it).

(E) For offices to house the manager, operating personnel, and programmers--400 square feet (convenient to the facility).

(F) For key-punching personnel--100 square feet (convenient to the facility in an acoustically controlled room).

The central processing facility itself needs to be environmentally controlled:

(A) Temperature held near 75 degrees F. (in the range of 60-70). (Since the heat load generated by a typical installation is about 52,000 BTU, this implies roughly 4 1/2 tons).

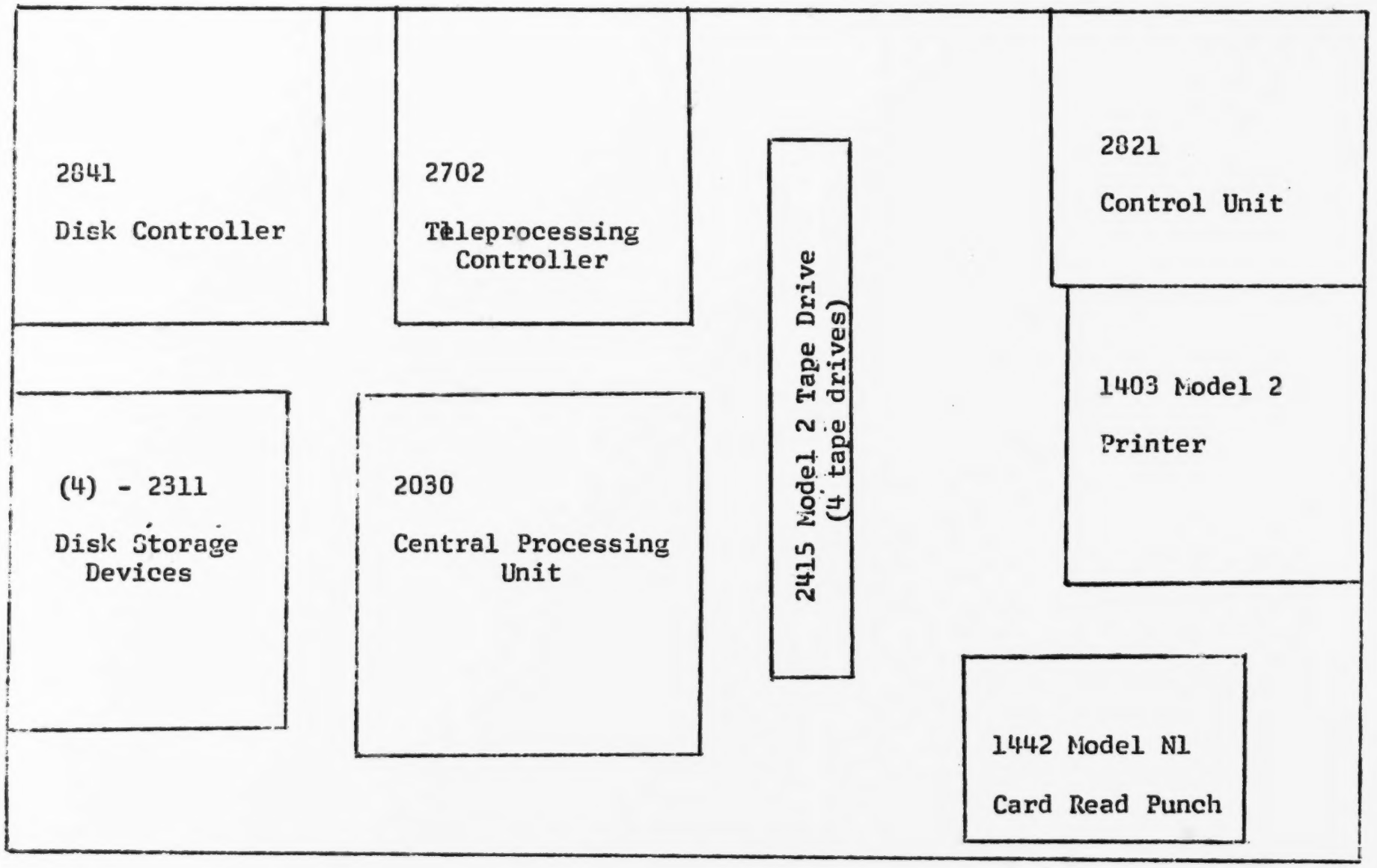
(B) Humidity held near 50% relative (in the range of 40% to 60%). This has particular significance for the magnetic tapes which ten to change their operating characteristics under excessively low or high hunidity.

(C) Dust must be controlled according to preseribed standards, again primarily because of its effects upon the reliability of magnetic reading and recording systems.

(D) There should be recorders for both temperature and humidity so that as variations occur they can be pinpointed in time.

(E) A cut-off for the air conditioning system should be provided within the facility.

The required power supply is a function of the specific configuration of equipment. A typical load is 20 KVA at 175 amps. This can be either 208 or 230 and includes both 3 phase and single phase. Because a stable power source is essential, surges must be controlled to within 5 to 10%. This implies an "isolation transformer" a power cut-off in the facility, and a ground wire ("green" wire) to a well-defined "building ground". There should be a continuous recording of the voltage.



Notes:

--Any pieces of equipment listed in Figure 3 not shown are included in allocated areas.
--The areas shown for pieces of equipment include space required for clearance and access.
--The area included in the space plan is 40 feet by 25 feet, or 1,000 square feet.

FIGURE 5: Library Computer Space Plan

The various mechanical units--printers, card readers and punches, key punches, etc.--tend to be noisy. Acoustical control is thus essential particularly if card equipment is involved.

The wiring and cabling within the facility should be placed under the floor, which implies either a false floor raised 12" or a recess of 12" under the floor.

Peripheral units, including point-of-action recorders, typewriter-type terminals, and displays, will be located throughout the library building. They usually involve connection by cables to multiplexing units or buffers and then by telephone line to a teleprocessing terminal in the facility. In particular, a typical peripheral unit is an IBM 1030 Data Collection System which accepts pre-punched cards (such as book cards and borrower cards) and transmits information from them to a key-punch or to an on-line computer. A second type of unit is a typewriter terminal. These can have the necessary buffer equipment directly associated with each and thus require connection only to a telephone line. A third type of unit is the cathode ray tube display, which requires a higher transmission rate and uses a separate multiplexing unit.

V. SOFTWARE FOR THE CIS SYSTEM

The dominating technical constraint on CIS software (computer programs) is the requirement for the ability to handle data from a variety of existing files. The processing, preparation, and output of the data once it has been selected and extracted is a relatively straightforward (although by no means trivial) task. The heart of the matter, therefore, is the ability to maintain, read, select, and extract data from files prepared by other organizations. As it now is, each data base has its own format, its own thesaurus, and its own package of "file management" programs which provide capability for maintenance and search. Each data base now requires a separate set of forms and procedures for utilization. With twenty data bases, each representing three or four files, the installation would be faced with the spectre of perhaps 500 different operating programs and procedures, few of which would be compatible with the library's operating system.

Therefore, how do we add data bases without proliferating programs to the point of virtual strangulation? The answer might lie in standardization, but that seems hardly likely, in view of the enormous variety of purposes served by the data bases to those who originate them. It might lie in conversion of the data bases to some standard format and structure for storage and processing by the library using them, but this also seems unlikely, in view of the sheer bulk

of data involved. It might lie in the use of generalized file management programs which can handle the variety of data bases and provide standardized services based on them.

The conclusions, based on the work done to date, are that custom programming for each data base is too lengthy, too costly, and too unresponsive to the needs of the Center and its users. On the other hand, translation or transliteration of files for use in some standard system is impractical because of the possible loss of meaningful information, the costs, the continual changing of formats, and the difficulty in processing. These points, when combined with the uncertainty of future data base formats and the changing nature of user requirements, all suggest as the solution the development of a generalized system appropriate the Center operations.

The design of the generalized system will be special purpose insofar as it reflects the special requirements of the Center for Information Services. Many recently developed generalized file management techniques, however, can form the basis for system design.

To use such generalized programs requires a careful description of each data base, both so the generalized programs can operate on it. and so the user can know what level of service he can call on. Usually, these programs provide a clearly distinguishable set of stages of processing, from fixed field, fixed format processing (the simplest and most efficient), to variable format processing, to text processing. Their relative efficiencies differ so radically that the prospective user must be well aware of precisely what data from a given data base can be effectively processed by a given level of program.

SPECIFICATIONS FOR CIS FILE MANAGEMENT SOFTWARE

For the purpose of clarity, this section describes the capabilities of the CIS File Management system as if it were operational, using the present tense, rather than future, throughout the section.

The CISFMS (Center for Information Services File Management Software) is a general purpose file management system. That is, a great variety of file structures may be defined independently of the processing functions performed.

It may be said that any computer programming language is general purpose in the sense that it is not limited to particular files and functions. In order to relieve the programmer of some detail, the notion of higher level languages was developed. The best known of these languages are COBOL, PL/1, FORTRAN, and ALGOL. The use of these languages is said to result in an average reduction of about 5 to 1 in the number of instructions which must be written by the programmer to perform a given application.

CISFMS introduces a still higher level of communication between the user and the computer. By relieving the user of many more requirements to communicate his needs to the computer, CISFMS permits use of the system without formal training in computer programming. Through the concept of different levels (subsets) of communication between the users and the computer, CISFMS may be used by library personnel, system analysts, or computer programming specialists--at the appropriate level of detail. Thus, instead of employing assembly language of a higher level language, the CISFMS user employs a small

set of structured forms to describe his problem solution in the amount of detail required.

CISFMS is used for producing computer programs for normal day-to-day operations, as well as for specialized requirements. The functions which may be involved in such operations include the creation and maintenance of files from original input (e.g., punched card and magnetic tape data), the selection of records from files according to either defined or computed criteria, computations involving data from selected records, extraction and sequencing of results dependent on these data, and the formation of new files for other, subsequent use. As we have said, the files(s) and the function to be are independent of each other, thus providing great flexibility in the use of CISFMS. In execution, however, they are tied together in order to minimize the information which must be provided by the user.

File Definition

CIS File Management Operation is centered around the concept of master files. In order to extract or retrieve data from files, the problem statement must refer to previously defined field names in specific files. When processing requests are presented, the files with which they deal therefore must have been previously defined.

The file definition specifies certain overall file parameters (such as record format and block size). More importantly, the record structure is described also.

CISFMS will have the capability of reading record structures which are fixed or variable in length and whichcan contain:

1. Variable length fields and segments.

2. Repeated fields and segments of the same type.

3. More than one type of format of field or segment at any hierarchical level.

4. An adequate number of hierarchical (nested) levels of segments within a record.

5. Various techniqaes to identify the format types and sizes of records, segments, and fields in a file.

File Organization Concepts

The organization of a file is generally independent of its specific content. Thus, files can be organized sequentially, in terms of some field in the data items in the file; randomly, so that records must be located by reference to an index or an algorithm; or in other ways.

File Search Concepts

The processing of CIS files must begin with the search of a particular file to select records for subsequent use by a requestor. The CISFMS provides capabilities for the simplest forms of such a selection. An obvious extreme is to provide the requestor with a copy of all records in a file. Normally, of course, more selective search criteria are specified. One may request, for example, records identified by particular data values in specified fields (e.g., specified document numbers or subjects). Still more complex search criteria may seek to relate a set of data values in each record to one another for the purpose of selecting those records in which specified relationships exist.

In general, CISFMS allows two file search approaches. In the first, all of the search criteria are specified by the requestor at the start of the search. Since record selection will depend on values actually contained in the records, either index records must be processed or the entire file must be searched to select all applicable records. The second search approach generates search criteria during the course of the search. In both cases, however, CISFMS handles requests phrased in terms of "field-structured" data.

The retrieval capabilities of CISFMS enable the users to select and extract data from the files. The key to effective retrieval is the logical selectivity of the system. CISFMS capabilities include an appropriate set of comparators, Boolean connectors, and types of comparands. Conditional expressions may be combined and a number of nesting levels is provided.

System Monitoring

CISFMS monitoring capabilities include provisions for:

1. Preparing utilization statistics by user, file, type of request, etc.

2. Cost accounting and bharging of accounts.

3. Protection of proprietary files.

System Functions

CISFMS is capable of performing many file management functions:

1. Read existing files from punched cards, magnetic tapes, and other machine-readable input.

2. Maintain files by making additions and deletions.

3. Reformat files to reflect changing specifications and requirements.

4. Select, from files, records that contain data of interest in a problem.

5. Extract data items from the selected records, or use whole records.

6. Arrange output by sorting, sequencing, and grouping.

7. Format printed reports that contain such elements as Preface, Page, Title, Page Number, Column Headings, Column Footings, Line Numbers, Detail Entries, Summaries, Statistics, Line Count, and other details that make a printed report or document informative and attractive.

8. Summarize data to as many levels of total and sub-totals as required, with wide flexibility in format and content of printed output.

9. Compute new values based on values in the file, for use in selection, further computation, printed output, subfiles, or the updated file.

10. Produce printed reports or other printed documents such as 3 x 5 cards, labels, or output on preprinted forms.

11. Produce subfiles on cards, magnetic tape, disk, or other media for further processing by CISS or other systems.

System Operation

The system will provide for the storage of source programs in a "library" for subsequent compilation. By storing the source program, rather than the object program, the system enables the user to conserve space in his system library for other purposes. In operation the user has the option of re-running such programs by recalling them either in source or object language form and operating under the system. This capability supplements the ability to define new data base requirements.

The capability to maintain and query master files, once the user has defined the master file and the query specifications, is then essentially automatic. This type of implicit specification is a basic design concept of the system. For example, a "standard" mode of operation

will automatically be invoked unless the user specifically requests an alternative mode. These standard cases are applicable in many situations.

The most important advantage of CISFMS is its simplicity of use. It makes use of "programming by questionnaire", in which the user merely answers a series of questions describing the results he requires. An ordinary search request can be described directly by the research or library-oriented user in a few minutes. More complex and sophisticated problems can be described to CISFMS in a few hours.

In summary, the Center for Information Services Software allows the library to use oomputers in the handling of many separate files with a minimum of lapsed time between acquisition of a data base and operational use of it. It reduces the demands for skilled programmers and analysts, and minimizes communications problems between the academic community and data processing people.

SPECIFICATIONS FOR CIS REFERENCE RETRIEVAL SOFTWARE

Since the basic CIS File Management Software provides capability only for the simplest, field structured search logic, the CIS system of software must also include a module for the processing of more complex requests. This is called the CIS Reference Retrieval Software (CISRRS), since it is of primary value in searching of reference data bases which involve the use of "subject" descriptions.

File Definition

The file definition for the CISRRS module is identical with that of the CISFMS module. Those fields of particular concern are the "repeated fields", which are characteristic in reference retrieval situations.

File Organization Concepts

The need for methods of organization beyond those of the sequential and indexed sequential, used in the CISFMS, is evident. A variety of indexing aids must be included:

1. "Inverted files (such as key-word indexes).
2. Dictionaries, hierarchically structured subject headings, and thesauri.
3. Word frequency lists and tables of statistical association.

The CISRRS must provide for the maintenance of these indexing aids as well as for the use of them in the formulation and processing of search requests.

File Search Concepts

Search in the CISRRS module differs from that in the CISFMS in at least two respects:

1. It involves simultaneous, interactive processing of at least two files (the master file containing the data of interest, and index files).
2. It provides more sophisticated processing of repeated field data.

In particular, search requests can be formulated as Boolean combinations of terms as well as of specified field values. The terms will be search for in the indexigd aids, and provision is made for automatic explosion of them based on theseet of inter-term references found. The CISRRS includes capability for correlating index records, based on defined request logic, to derive master file entry references for subsequent processing.

System Monitoring

CISRRS includes, in addition to the monitoring functions of CISFMS, the Maintenance of statistics on inter-file reference.

System Functions

CISRRS supplements the file management functions of CISFMS by its ability to:

1. Maintain index aids from master file records.
2. Explode request terms based upon data stored in master files and index files.
3. Correlate data from separate index records or master file records.
4. Search two or more files simultaneously.

SPECIFICATIONS FOR CIS TEXT PROCESSING SOFTWARE

Although in principle either the CISFMS or the CISRRS could process text data by treating each word as a separate entry in a repeating field, such processing is relatively inefficient. To provide specific functional capabilities, the CIS system includes a module, called the CIS Text Processing Software (CISTPS), designed around the particular needs in generalized text data processing.

File Definition

The file definition for the CISTPS module is identical with that for the CISFMS-CISRRS modules. Those fields of particular concern are the "text fields". Particular attention must be given to provide for "character coding" of multiple font text.

File Organization Concepts

The CISTPS uses the same kinds of indexing aids involved in the CISRRS. However, their scope of coverage is likely to be much broader,

since all terms appearing in text, must be considered (as terms either to be processed or not to be processed).

File Search Concepts

Although search logic considerably more complex than that provided in CISRRS appears to be desirable (including, for example, automatic parsing), it is not possible to specify at this time an adequate, operational definition of it. Therefore, the search logic of CISTPS is identical with that of CISRRS.

System Functions

CISTPS supplements the file management and search functions of CISFMS and CISRRS by its ability to:

1. Produce concordances and other word lists.
2. Collate texts for the detection of differences and similarities.
3. Accumulate statistics on frequency of occurrence of words and word strings.
4. Derive indexing terms based on a variety of clues, including frequency of occurrence, format, context, etc.